DOCUMENT RESUME

ED 087 786 TM 003 394

TITLE	CIRCUS: Comprehensive Assessment in Nursery School and Kindergarten.
INSTITUTION	Educational Testing Service, Princeton, N.J.
PUB DATE	21 Aug 73
NOTE	41p.; Proceedings of a symposium presented at the American Psychological Association Convention (Montreal, Quebec, August 21, 1973)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Child Development; Divergent Thinking; *Early Childhood Education; Educational Development; Individual Development; Kindergarten Children; Language Development; *Measurement Techniques; Memory; Perceptual Development; Preschool Children; *Preschool Tests; Problem Solving
IDENTIFIERS	Circus Tests

ABSTRACT

These papers deal with four specific propositions concerning the role of measurement in early childhood education: 1. measurement should play an integral part in early education, independent of special pressures to evaluate program effects; 2. the measures should be designed or adapted specifically to the continuing needs of preprimary educators and to the limitations in time and measurement expertise typical of many nursery schools and kindergartens; 3. there should be no lowering of technical standards for the instruments which assess young children; 4. theoretical bases and construct validity are just as important for measures intended for use in practical settings as for research instruments. Six speakers at the symposium explained their reasoning behind these propositions and illustrated their remarks with descriptions of CIRCUS, a program of new instruments and supporting services for preschool and kindergarten teachers. Titles of the six presentations are: Assessment for Personal and Educational Development; Language Comprehension and Performance; Memory and Experience; Quantitative and Relational Understanding; Problem Solving and Divergent Production; and, The Context of Assessment and the Assessment of Context. Two additional speakers presented their critical views of these measurement approaches. (Author/NE)

TM 000 294 ED 087786

CIRCUS

Comprehensive Assessment in Nursery School and Kindergarten

Proceedings of a symposium presented at the American Psychological Association Convention, Montreal, Quebec August 21, 1973

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

Dorothy Urban

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

Educational Testing Service is an Equal Opportunity Employer.
Copyright © 1973 by Educational Testing Service. All rights reserved.

Library of Congress Catalog Card Number: 73-90524

CONTENTS

iv Foreword, Esther Kresh, Office of Child Development, Department of Health, Education, and Welfare

7 Assessment for Personal and Educational Development, Scarvia B. Anderson, Educational Testing Service

12 Language Comprehension and Performance, Masako N. Tanaka, Far West Laboratory for Educational Research and Development, and Carolyn E. Massad, Educational Testing Service

19 Memory and Experience, Gerry Ann Bogatz, Educational Testing Service

23 Quantitative and Relational Understanding; Perceptual Skills, Ann Jungeblut, Educational Testing Service

27 Problem Solving and Divergent Production, Ruth B. Ekstrom, Educational Testing Service

33 The Context of Assessment and the Assessment of Context, Samuel Messick, Educational Testing Service

39 Discussion, Boyd R. McCandless, Emory University

40 Discussion, Marshall S. Smith, National Institute of Education

FOREWORD

Everybody seems to be talking about early childhood education these days and everybody seems to be doing something about it. This is particularly true in the field of measurement. In the Head Start Test Collection, for example, there are between 3,000 and 3,500 different instruments for measuring young children's behavior. These tests cover a great deal of the cognitive and affective domain (not to mention office space) and range widely in their usefulness. As Henry Dyer, for many years a Vice President of Educational Testing Service, remarked in a talk some years ago: "Some are psychometrically respectable; some are trying to become respectable; and some are innocent of any psychometric properties whatever." Many, too, are instruments whose focus is too narrow to be of use to teachers or of such complexity that even psychologists cannot administer them without special training.

Last summer, a number of people who are concerned about this problem took part in a symposium at the American Psychological Association convention in Montreal. Called "CIRCUS: Comprehensive Assessment in Nursery School and Kindergarten," the symposium was addressed to the propositions that:

1. measurement can and should play an integral part in early education, independent of special pressures to evaluate program effects;

2. the measures should be designed or adapted specifically to the continuing needs of preprimary educators and to the limitations in time and measurement expertise typical of many nursery schools and kindergartens;

3. although assessment of young children presents some special challenges to traditional psychometric criteria, there should be no lowering of technical standards for the measuring instruments;

4. theoretical bases and construct validity are just as important for measures intended for use in practical settings as for research instruments.

Six speakers at the symposium explained their reasoning behind these propositions and illustrated their remarks with descriptions of CIRCUS, a program of new instruments and supporting services for preschool and kindergarten teachers. Dr. Boyd McCandless of Emory University and Dr. Marshall Smith of the National Institute of Education presented their critical views of these measurement approaches. Their papers appear in the following pages.

Esther Kresh
Office of Child Development,
Chairman

ASSESSMENT FOR PERSONAL AND EDUCATIONAL DEVELOPMENT

Scarvia B. Anderson
Educational Testing Service

Measurement has loomed large in recent attempts to evaluate the effects of Head Start, Follow Through, Sesame Street, Title I, and other innovative preprimary programs. The zeal with which researchers have developed new instruments to use in these efforts has been matched only by the enthusiasm with which they have grabbed older measures off the shelves. The evaluation results have given social scientists new, if sometimes conflicting, insights into children's educational and psychological development and have been the occasion variously for rejoicing and despair by program sponsors.

The evaluation results have affected individual teachers, children, and parents little if at all. This is not a surprise—or even a criticism—because these large-scale studies of children and programs en masse have generally not been designed for that purpose. However, it is important that comparable attention be given to the uses of measurement to further the *development of individuals*. Specifically, we need measures designed to diagnose children's educational needs and to aid teachers in selecting appropriate classroom strategies to meet those needs.

In calling for such measures, we recognize that the literature of tests and measurements is not without some ingenious instruments and measurement approaches applicable to the instructional needs of children. Such developments, however, have tended to be a minor theme in a field dominated by:

- global assessment of generalized traits such as intelligence, neuroticism, or reading ability; emphasis on measurement of one such trait at a time; the use of measures primarily in the service of institutional needs (such as selection and evaluation) rather than the needs of individuals

- preoccupation with measurement of people, with little comparable attention to assessment of the environments from which, and in which, they are expected to function

- standardization, rather than ratiocination, as the chief means of establishing criteria of "adequate" test performance; assignment of a

primary role to predictive validity, as opposed to content- or construct-validity

- concern with measuring maximal rather than typical performance; preferences for treating most educational and psychological variables as linear and continuous, where "more" is "better"

- application of formal measurement instruments primarily to populations of older children and young adults to whom test taking has become an expected part of their lives

In the remainder of this paper I shall deal in turn with the special problems that each of these conventional emphases presents to those of us attempting to assess young children in the interests of their personal and educational development.

Differentiated vs. Global Assessment

Much has been written in recent years decrying the uncritical use of IQ measures and espousing a differentiated view of human capability. Less has been said about the implications of such stands for teachers and young children. Since it is true that tests can influence both how teachers view children and what they do about children, measures based on a *differentiated* view of the child can help reinforce a teacher's understanding of the complexity of the child and the broad range of skills, achievements, coping styles, and other factors that characterize his development. Measures that emphasize a child's potential for progress and improvement can influence the educational decisions the teacher makes and the treatments she applies in a way that is very different from measures of "fixed traits." For example, IQ scores have tended to provide teachers with excuses: "Henry only has an IQ of 85 so what can I do?" But information that "Henry can classify objects according to one but not two attributes" presents a challenge to her instructional talents.

The Effects of Environment

The use of tests to select college freshmen or industrial employees is a relatively pragmatic endeavor. At least until recently, an admissions or personnel officer was little concerned with *how* a person came to have the characteristics revealed by the tests. But this is not true of teachers.

They need to know as much as they can about the environmental factors that might influence the child's educational development. Educators are especially concerned with the effects of educational environments on children. At a minimum, we simply cannot interpret a child's score on a measure and use it as a basis for instructional prescription unless we know something of the context in which that score was obtained, including the kind of educational treatment the child has been exposed to previously.

Norms and Other References

The dependency of the tests and measurements profession on normative interpretations of test scores is at least partially responsible for the current popularity of the so-called criterion-referenced tests. We think it is unfortunate that objectives-referenced, content- or domain-referenced, and construct-referenced measures are all lumped under the "criterion-referenced" label, for that obscures the fine and important distinctions among them (see "Criterion-referenced measurement" in Anderson *et al.*). However, it is even more unfortunate when a sharp contradistinction is drawn between all these and the norms-referenced tests, especially when there is an accompanying connotation of "good" versus "bad" measurement practice. The two approaches can usefully supplement one another; this is especially apparent in the child development area where age- as well as stage-development scales are important. Furthermore, normative notions, although frequently implicit, underlie or circumscribe criterion-referenced measurement. For example, we do not ordinarily try to measure a five-year-old's ability to comprehend a *New York Times* editorial.

The Properties of Variables

The technical problems involved in the reliable and valid assessment of young children cannot be denied. Neither, however, are they insurmountable. Development begins, as all measurement attempts should, with appropriate conceptualization of the traits and domains of interest and should include specific attention both to the nature of the measurement objectives and the properties of the variables involved. We mentioned earlier that, historically, educational measurement has been primarily concerned with documenting maximal performance. Typical performance is not so easy to assess reliably and that is probably why experts have tended to shy away from it. However, we have some

evidence that measurement through unobtrusive observations in natural settings stands a better chance of providing a basis for inferences about typical performance than do formal test-like encounters.

It is easy to go along with the notion that knowing more letters (or colors or numbers) is better than knowing fewer, and it makes sense to give the child who knows more letters a higher score. However, such variables as response latency are not so easily interpreted; while quick responses may indicate lack of reflection, very slow ones may be more indicative of obsessiveness or fatigue than of reflectivity. We know, too, that some dimensions may be bipolar, and extreme behavior at *either* end may be maladaptive. (The attempts to assess "self-concept" have suffered from failure to take account of such possibilities.) The fact that different variables may show different developmental trends is relevant here, too. For example, some abilities may increase with age and training (perhaps tapering off at later ages or with lack of practice), while others may decrease with maturity, or be cyclical, or remain fairly constant across wide age spans.

The Young Child as Examinee

The ultimate trick, of course—as with subjects of any age—is to ensure that the tester's task becomes the child's task, or that the tester's interpretation of the child's responses corresponds to the real meaning of those responses. More measurement efforts may have foundered here than at the conceptual level. The problem may be one of finding a meaningful response that a child is capable of making to a stimulus (for example, investigators have harnessed the orientation reflex in studies of attention processes in infants). Or it may be one of eliminating irrelevant difficulties; how many of us have tended to draw conclusions about auditory discrimination on the basis of a measure that, at the younger ages at least, was more a test of the child's understanding of the concepts of "same" and "different"? We must be concerned, too, with the special problems of obtaining responses from minority/poverty or handicapped children, the limitations on time and clinical expertise of those typically charged with assessment in early education programs, and the difficulties of applying traditional psychometric theories and principles to instruments and populations that are nontraditional. For example, "guessing" would seem to have a very different meaning for the high school student taking the *Scholastic Aptitude Test* than for a five-year-old identifying pictures related to "real world" sounds. Similarly, we expect sophisticated test takers to recognize that in a

multiple-choice test the correct answer is likely to appear in any of the response positions; however, young children faced with difficult items may be more likely to respond in terms of position biases or other types of response sets (Anderson, Messick, & Hartshorne).

* * *

Earlier this year, a group of child development experts helped us delineate 29 aspects of social competency in young children (Anderson & Messick, 1973). The CIRCUS instruments we will refer to in these papers as examples of strategies for assessing young children in the interests of their personal and educational development tap aspects of only 13 or so of these. Nevertheless, they represent a start toward comprehensive assessment of both children and their educational environments. Almost more important, they seem to live up to the promise of their name by being fun for both teachers and four- and five-year-olds.

References

Anderson, S., & Messick, S. Social competency in young children. *Developmental Psychology*, in press.

Anderson, S., Messick, S., & Hartshorne, N. Priorities and directions for research and development related to measurement of young children. Princeton, N.J.: Educational Testing Service, October 1972, PR-72-22.

Criterion-referenced measurement. In S. Anderson, S. Ball, R. Murphy, & E. Rosenthal, *Anatomy of evaluation*. Washington, D.C.: Office of Naval Research technical report on Contract No. N0014-72-C-0433, 1973.

LANGUAGE COMPREHENSION AND PERFORMANCE

Masako N. Tanaka
Far West Laboratory for
Educational Research and Development

Carolyn E. Massad
Educational Testing Service

It has been suggested that one of the purposes of assessment is to "make the child visible." In the study of language comprehension and performance in very young children, the problem of how to increase visibility requires us to both widen our angle of vision and to sharpen our focus of view. We need not only to look at a larger variety of behaviors but also to obtain enough instances of a particular behavior so that it can be seen clearly. At the same time, it is important to develop this visibility in ways that would be helpful to those working with children in an educational setting. The selection of the particular ways of looking at child language thus should be based on those elements that can be assessed in the usual classroom context and that have some basis in the research literature as being important in the development of language in children.

Evolution of the CIRCUS Language Measures

Much of the work in the development of the language measures used in the CIRCUS collection was based on prior experience with similar measures in various research studies conducted at Educational Testing Service by the Early Education Group and the Head Start Longitudinal Project. These earlier measures, in turn, incorporated and adapted a number of ideas and item types used by other researchers to whom we are greatly indebted.

For example, within the theoretical context of looking at language development, Carroll (1964) has suggested that there are two main classes of functions of language:

1. as a system of responses by which individuals communicate with each other (inter-individual communication) and

2. as a system of responses that facilitates thinking and action for the individual (intra-individual communication).

As part of an earlier ETS study, Shipman and Bussis (1968) suggested that these two functions could be identified from a linguistic point of view, and that different word classes included in the grammatical structures of the child's speech could be identified with these functions. In their analysis, the group of words called *content* words is primarily used for communication between people, whereas the group of words called *functor* words is responsible for facilitating thinking. The content words (such as nouns, verbs, and adjectives) carry most of the communication load: When a child says "Mommy, cookie," her mother knows she means "Mommy, please give me a cookie." The functor class of words, which comprises only about one percent of the total vocabulary, consists of auxiliaries, prepositions, articles, pronouns, conjunctions, and inflections. Although functor words convey little information in and of themselves, they make a critical difference in meaning when used in context.

This research interest in the development of words in both content and functor classes is represented in the CIRCUS language instruments. The use of content words is primarily studied through the use of single-word measures such as a picture vocabulary test (*What Words Mean*) and an auditory discrimination test (*How Words Sound*), whereas the use of functor words is studied by three different measures: *Listen to the Story*, a listening comprehension test, *How Words Work*, which measures the receptive understanding of certain grammatical constructions, and *Say and Tell*, which measures the ability of the child to produce the same or similar constructions. These last two measures are designed to provide information that can be used to compare the child's *receptive* vs. *productive* use of grammatical structures.

The Purpose of the Language Measures

The purpose of the CIRCUS measures is to provide the teacher of the four- and five-year-old child with a *reasonable* sampling of the child's language. The word *reasonable* is used quite deliberately, and it applies in a number of different contexts. We must all agree that the best sample of language in terms of range, content, and adequacy would be that obtained by the continued and careful observation of the child by a sensitive observer over a long period of time. A reasonable sampling must, however, be limited to that which can be done by a relatively untrained observer in an appropriate period of time under realistic classroom conditions. We would also agree that there are a number of research directions that are provocative in terms of developing an

understanding of the child's language, but a reasonable approach would be to select those that appear to be most closely related to the educational goals of the teacher in the classroom.

The use of the word *reasonable* in a more positive sense requires that we provide as large and adequate a sampling of the child's language as is possible under the constraints of a standardized assessment situation. That is, in contrast to many so-called readiness measures, it is our feeling that if a particular language behavior is important enough to measure, there should be enough instances of that behavior so that one can look at it carefully. For example, if the ability to listen is an important area to observe, then there should be more than one way to assess it, and the number of items on each type of listening behavior should be sufficient so that the teacher obtains some instructional information from an analysis of the items.

Increasing the Amount and Kinds of Feedback

The growth of listening skills may be considered as the construction of a sound-symbol system in which the spoken word is associated with a representation, either internalized (imagery) or externalized (object or picture). In the CIRCUS instruments, the development of this system is monitored through the use of separate measures that assess various abilities such as connecting sounds with pictures (a child recognizes a picture of a bell upon hearing the sound of a bell on tape), discriminating sounds within words (auditory discrimination), understanding words connected together as in stories (listening comprehension), and coping with the linguistic use of language (use of inflections, prepositions, pronouns, and so on). Thus, instead of a global score on listening comprehension or a readiness score based on a collection of a few items from each of the above categories, the teacher is provided with specific information that would be useful in an instructional program. That is, instead of finding that half of a class is "not ready" for reading, the teacher has some indication of the *kinds of items* that are difficult for a particular child or group of children.

In addition to increasing the *amount* of informational feedback to the teacher, the development of items also has involved a concern for the *kind* of feedback available. For example, teachers might make more productive use of the wrong answers given by children. Whenever possible, therefore, the distractors in the test items have been carefully designed so that the teacher can analyze the wrong answers to help her plan her instructional program. If an item requires the use of several

elements to be correct, such as "Clarence Clown has a big nose and a smiling mouth. Mark Clarence.," the distractors have different elements that are incorrect (a clown with a big nose and frowning mouth or a clown with a smiling mouth and a little nose). Again, if an item requires the child to attend to a sequence of directions, the design of the distractors helps a teacher to see whether the child consistently tends to listen to either the beginning or the ending part of a phrase. This philosophy of a testing/teaching approach to test development has had an earlier history at ETS with the ETS *Cooperative Primary Tests.* Providing teachers with ways in which to use such information as part of their instructional program has been a very rewarding experience.

Picture Vocabulary Tests: A Hazard

Space does not permit a full discussion of each of these listening measures, but the use of picture vocabulary tests is so common that it warrants some mention here (and thoughtful consideration on the part of test developers). Perhaps more than any other type of measure, the assessment of the child's vocabulary through the use of pictures must be viewed as a hazardous undertaking. If we agree that words are symbols or abstractions representing concepts, we see that the use of a picture vocabulary test incorporates the folly of trying to measure the concept of a class or category with a single instance of that category. In other words, we are trying to measure whether a child understands the concept of "dog" with a picture of a single, particular dog. In a sense, this procedure violates the developmental notion of label acquisitions in which we assume that the child learns to abstract the concept of "dog" from a variety of instances. That is, that the wider the representation of instances (the number of kinds of dogs), the broader and more generalizable is the child's concept of "dog." The assumption of the picture vocabulary test is that the child chooses the correct drawing as a categorical response. The hazards of this assumption are clear: One child may get the correct answer simply because the pictured dog closely approximates the *only* dog he knows rather than because he knows a large number of dogs and is able to generalize to the class in question.

The future development of picture vocabulary tests should be concerned with some resolution of this problem. One approach may be to provide as many "drawable" examples of the target word as possible. The child's task would then be to identify these examples out of a set of nonexemplars. Such a procedure would provide information on the

breadth of the child's knowledge of a particular word rather than on whether he happens to recognize one specific version. For the present, however, our work with the CIRCUS vocabulary measure represents an attempt to correct a problem that is common in many of the picture vocabulary tests used for this age range. Quite often, the items in such tests measure only the child's global understanding of a word. Thus, the distractors have little or no relationship to the target word, and the child needs only a vague association with the required word in order to eliminate the wrong answers. In the development of the items in the CIRCUS vocabulary test, there was a deliberate focus on the careful use of distractors that would measure the preciseness of the child's understanding—if the stimulus word was "log," the item included drawings of a piece of lumber and a tree as well as a log.

The Real World of Language Development

In contrast to measures that focus on receptive language, the real world of language development is to be found by listening to the productive speech of children. If we were to walk into a room full of four- or five-year-old children, our main impression would be an awareness of the hum of children's voices. There is a tremendous amount of talking going on, some of which may be elicited by the adult but much of which is spontaneous. Here, then, is the real world of oral language in the young child. This is where he learns to use language to deal with his world in all its complexity—to ask questions, to get help, to imitate, to role play, to order other children around, to say, "Hey! Look at me!"

We agree that this real world of language performance cannot possibly be fully explored through the use of any prescribed set of standardized measures. At the same time, there is a need to provide some way of helping the teacher to sample the richness of the child's oral language. *Say and Tell* measures the growth of the child's spoken language by observing three types of language use:

1. *The descriptive use of language:* The child is handed a common object and is asked to describe it. One item elicits the child's use of categorical language such as asking for various attributes ("What color is it?"). Another merely asks him to "Tell me all about that."

2. *The functional use of language:* The child is shown a number of pairs of drawings. A statement is made about one of the pictures,

and the child is asked to complete the statement that applies to the other picture ("Here is a boat. Here are two _____".) There are more than 40 items dealing with such things as the use of plurals, verb tenses, prepositions, subject-verb agreement, comparatives, possessives, and so on.

3. *The narrative use of language:* The child is shown a large colored drawing, and the teacher explains that it is a picture out of a storybook, but that "I don't have the story that was in the book, so I want you to make up a story to go with this picture. What do you think the story was about?"

There are two items, and the child's story for each picture is taken down verbatim. Each story is scored for both quantitative and qualitative dimensions. The quantitative scoring includes the more traditional measures of the number of words and the number of different words. The qualitative scoring measures the "storyness," or the use of elements such as action, imagery, effect, characterization, and organization. It is unfortunate that the use of written protocol prohibits observation of some of the richest elements of the child's oral language. Much of the effectivness of a young child's communication is apparent in his use of such elements as intonation, pacing, and volume (loudness), as well as the important nonvocal elements of facial expression, gesture, and body language. However, it is our hope that by providing the teacher with information on the qualitative elements of the written version of a child's story, she will become more aware of the complexity of the child's use of language for communication.

A number of other researchers have focused on the comparison between the child's receptive vs. productive use of language and have found that the child can understand a much larger number of words than he can use in his own speech. In contrast to receptive language measures, which require a child to select from a limited number of responses, the measurement of productive language is complicated by the fact that the variety of responses is limited only to the extent of the child's oral vocabulary and ingenuity. The authors' research with the Story Sequence Task in the ETS Head Start Longitudinal Study has supplied additional evidence that the young child is quite capable of understanding the meaning of a word used in a story although he cannot recall the exact word in his retelling of it. For example, one of the stories included a statement that Mr. Turtle *visited* his friend Mr. Pig. In the subsequent coding of the children's version of the statement,

we found that there were some eight or nine acceptable ways in which the meaning of the word "visited" was communicated: "he went over to get," "he asked him to come over," "he went to play with," and so on.

This same type of ability to understand the *intent* of a communication combined with an ingenuity in the use of the child's own language is also apparent in the children's response to the CIRCUS productive language measures. In the measure of functional use of language, many of the responses showed that the children clearly understood the task but were managing it in their own language. For example, in one of the items on verb tenses, the teacher pointed to each of two drawings of monkeys and said, "*This* monkey ate his banana. *This* monkey is still _____." Back came responses such as, "This monkey is still *not finished*," "This monkey is still *hungry*," "This monkey is still *chewing*," "This monkey is still *holding his banana.*" As a result of this delightful but frustrating experience, we now have a tremendous respect both for the young child's command of his language and for the coding problems of researchers who have been working in this field.

Toward More Visibility

The development of language measures that provide as much visibility as possible is particularly critical today because many educational decisions about children are based on competency in language. The measures discussed in this paper represent our attempt to translate the current state of the art of language assessment into instruments that will contribute to that visibility by providing useful information to educators and researchers working with young children.

References

Brown, R., & Bellugi, U. Three processes in the syntax. *Harvard Educational Review*, 1964, 34, 133-151.

Carroll, J. B. *Language and thought*. Englewood Cliffs, N.J.: Prentice-Hall, 1964.

Shipman, V., & Bussis, A. Further considerations concerning the measurement of verbal behavior. *Disadvantaged Children and Their First School Experiences*. Princeton, N.J.: Educational Testing Service, December 1968. PR-68-4, pp. 67-68.

MEMORY AND EXPERIENCE

Gerry Ann Bogatz
Educational Testing Service

Memory is inextricably tied to experience, since one can only remember what one has experienced. But since no children have had identical preschool experiences, a test of memory must be based on experiences that are new and shared by everyone in the classroom. A child's long-term experiences and his ability to recall them can be assessed by a test of general information. Both these dimensions are measured by CIRCUS instruments described in this paper.

Measuring Memory

Memory is an important element in cognitive behavior, involving an ability that is associated with other cognitive processes such as evaluating or transferring ideas. A young child must not only remember what he has just been told in order to follow directions but must also be able to understand and remember things he has learned over time so that he can use them in appropriate situations in the future. Memory ability is complex, combining the abilities to attend to a stimulus, to retain or store it in some organized way, and to recall or retrieve it when needed. In addition, the ability to remember is greatly influenced by a wide range of variables, including the amount of intervening time among attention, storage, and retrieval, the kinds and amount of interference before retrieval, and the importance the child attaches to the material to be remembered.

No assessment of memory can encompass all these facets. However, even a partial assessment is important, since a child's ability to remember will affect almost all of his other cognitive and social experiences and performances. A teacher's recognition of the limits of a child's memory is essential for the sequencing of all learning activities (How long and how complex can an instructional sequence be and still be remembered by a child?); for the pacing of instruction (How much unrelated activity should occur between learning activities that are important for the child to remember?).

See and Remember, the memory task in CIRCUS, is an assessment of a child's visual memory and his ability to retrieve visual stimuli. It includes assessments of immediate recall (after a few seconds) and of delayed recall (after a few minutes and after intervening stimuli). As

well, it assesses the child's ability to remember a single stimulus ("Which one of these did you just see?"), the serial order of objects (train, cage, cart; cage, train, cart; cage, cart, train), the position of an object (ball over one of two seals rather than between them or over the other), and paired associates (associating proper names attached to animals).

See and Remember is designed to optimize the child's performance in the following ways:

- The child is told that he will be asked to remember each stimulus as it is shown, thus increasing the likelihood that he will attend to the material.

- The child is given several seconds to look at each stimulus so that he will have the opportunity to impose some structure or meaning on it.

- The child is allowed to practice with the task, and the items progress from the simplest to the most difficult so that experience with both the materials and the task is a built-in part of the measure.

- The child is asked to remember things that are familiar to him (animals), since things that are most familiar to a person are easiest to remember.

Memory in *See and Remember* is thus measured by a child's retrieval of a variety of visual stimuli both immediately and also over some time and intervening stimuli. The assessment of various memory skills within the one measure is intended to help the teacher determine whether they are difficulties related to the attention process, the storage process, and/or the retrieval process.

Measuring Experience

Measures of a child's general knowledge are often misinterpreted or even labeled "general readiness tests." However, the selection of items in a general knowledge measure should make it clear that what is being assessed is not anything as broad as "general readiness" but simply one aspect of a child's competency, specifically the child's accumulation of facts and concepts that are important to his functioning in school and at home. Whereas the other measures of CIRCUS assess specific

language, quantitative, perceptual, and problem-solving skills, the general knowledge measure assesses to some degree the extent to which the child has used these other skills to learn about things in his environment. In addition, it assesses the extent to which the child's environment and experiences may have failed to present him with some of this information.

Do you know. . . ?, the CIRCUS test of general information, includes measures of:

health and safety:	what's safe and never safe to handle; what foods do our bodies need
physical and social environment:	glass breaks; apples grow on trees; a man is older than a boy or baby
consumer concepts:	which thing costs the most money
music and literature:	who plays in a band; who surprised Goldilocks
TV and recreation:	what checkers look like; where Oscar on *Sesame Street* lives
practical arts:	what tells how hot it is; what is used to sew buttons

There are several reasons why a measure of general information is useful in an assessment package for young children:

- Children and adults are often judged by others on the basis of how much they seem to know. Most people would probably even go so far as to define intelligence in these terms (so-called intelligence tests are laden with knowledge and information items). In CIRCUS, however, the acquisition of general knowledge is clearly identified as a separate factor in the child's life by the inclusion of a separate general information test as one of many measures.

- A good deal of school time is spent teaching the child bits and pieces of information. At the preschool and kindergarten levels it seems appropriate to ask how much of this knowledge the child has acquired, especially if the knowledge assessed is geared to those bits

of information the child can be expected to need and use when he functions in the classroom and at home.

- Certain facts and pieces of information are assumed to be part of the knowledge of all people, even young children. Teachers and children alike are assumed to share a core of knowledge without which there could be no communication or understanding. Children who do not share this core may have difficulty functioning in school. An assessment of a sample of this core of knowledge can alert teachers to this need.

- The core of knowledge that is generally assumed to be a part of everyone's experience is likewise used, often unknowingly, to teach new information. Obviously, the further along a child goes in school without such basic information, the further behind he will fall in the acquisition of new knowledge. Immediate assessment of general knowledge is therefore needed to give an estimate of the child's basic core of knowledge.

- General knowledge is, in a sense, subject-matter specific. And the information a child has may or may not be related to various skills and styles he or she has acquired. It is certainly reasonable to ask how much information the child has acquired, and whether or not the amount of information is related to various skills the child has developed.

The measurement of memory and the measurement of experience both involve assessments of the storage and retrieval of information, the first dealing with immediate retrieval and the latter with long-term retrieval. Both processes are fundamental to an understanding of a child's intellectual development.

Reference

Educational Testing Service. Theoretical considerations and measurement strategies. *Disadvantaged Children and Their First School Experiences.* Princeton, N.J.: December 1968. PR68-4.

QUANTITATIVE AND RELATIONAL UNDERSTANDING; PERCEPTUAL SKILLS

Ann Jungeblut
Educational Testing Service

Our primary concern was to work within a theoretical structure of child development and within the constraints that we had set for ourselves in the conceptual development of CIRCUS as an array of preschool measures: The particular characteristics assessed must facilitate diagnosis of educational needs and selection of appropriate classroom strategies to meet those needs. The measures must be cast in a format that allows easy administration to small groups of children by classroom teachers so that the assessment environment becomes an integral part of the typical classroom routine. The CIRCUS theme should be pervasive since it provides familiar subject matter for both boys and girls, in all regions of the country, at all socioeconomic levels, and for all ethnic groups, thus minimizing assessment biases. And, above all, the tasks should be fun for, and intriguing to, young children.

Quantitative and Relational Understandings

Since our focus was on the age range from about 4 1/2 to 5 1/2, in Piagetian terms we are dealing, except in rare instances, with the non-conserving or preoperational child. In attempting to measure quantitative understandings, we were limited to developing group techniques to assess relatively global notions. For example, in the CIRCUS measure *How Much and How Many* the child is asked to mark among three pictures of elephants the "elephant that is largest," from among three pictures of ponies the "pony that is smallest," the "fewest seals," the acrobat with the "short pole," the "clown with the long nose," and to demonstrate his understanding of *most* in the sense of numerosity (which clown has the most balloons) and quantity (which cone has the most ice cream). It is these global notions, according to Piaget, that are the precursors of numerical comparison.

The work of Piaget and Inhelder (1969) has indicated that during the later stages of the preoperational phase, the young child understands the expressions and vocabulary of the next level although he rarely uses them spontaneously. For example, the young child who does not yet possess the least notion of conservation will describe pairs of objects in the following way: "That one has a big hat" (not "He has the bigg*est*

hat."), "that one has a small hat" (not "He has the smallest hat."), or "this one has a lot" (*not* most) and "that one has a little" (*not* least). Or, one member of a pair is described along one dimension ("This one is big.") while the second member is described along another ("This one is skinny."). Nevertheless, although he may not spontaneously use the expressions of the next higher level, he frequently understands them and can often select the "biggest" from among three objects. Comparison of receptive and productive skills, which is possible with a comprehensive array of measures like CIRCUS, will help to determine which level or levels a preschool child has attained. For example, the child who can select the *largest* elephant (from among three pictured in the *How Much and How Many* measure) and who can also come up with the word *biggest* (as required to complete the series *big, bigger, biggest* on the productive language measure *Say and Tell*) is very different from the child who correctly identifies the largest elephant but who completes the productive series by saying "the very big one" rather than producing the anticipated comparative form.

We were also concerned with the child's understanding of such notions of inclusion and exclusion as *some* ("Which picture shows some of the monkies riding?"), *all* ("Which picture shows all of the monkies eating bananas?"), and *none* ("Which picture shows none of the tigers jumping through hoops?"). As is obvious from the preoperational child's drawings, there seems to be no real awareness of perspective, but there is some notion of topological relationship. For this reason, it is important to assess such relational understandings as *on* ("Which picture shows the dog on top of the ball?"), *between* ("the clown between two elephants"), and *bottom* ("the clown at the bottom of the ladder").

The preschool child may well be able to recite the number names from one through ten, or even higher. However, at the preoperational stage, numerical evaluation is still linked with spatial arrangement, and the child may not recognize the need to match the number names in one-to-one correspondence to each object counted. Although the child need not conform to any specific order in counting, he must understand that each object must be attended to in turn and that he must somehow keep track of what has been enumerated and what remains to be enumerated. In terms of format, there is a need to use various combinations of stimulus and response modes—verbal stimulus/numeral response, and numeral stimulus/pictorial response. In addition to assessing these random correspondences that lead to number, in Piagetian terms, it is important to assess one-to-one correspondence in

the context of selecting the picture from among three that shows, for example, "just one ice cream cone for each of the clowns." In this case, the configuration of the two sets of objects can be identical or inverted.

The child's understanding of ordination is also of interest. It seemed appropriate to assess understanding of temporal order ("These pictures tell a story. Mark the one that shows what happens first.") in the problem-solving, or *Think it Through*, measure. On the other hand, spatial ordination seemed quite appropriate to include as an emerging quantitative understanding ("Here are some children waiting to get into the circus. Mark the last child."). It is also important to include the beginning concepts of quantitative negation—"Which picture shows fewer—not as many—dogs as the other?"

If we are to understand how the preschool child's quantitative development is related to, and affected by, his verbal development, the assessment of concepts such as those mentioned above is essential. Through an assessment array such as CIRCUS, we can identify for teachers the embryonic forms of quantitative concepts developing in the child and the reorganization necessary to move from one developmental level to the next higher one that are important to later mathematical development.

Perceptual Skills

It is admittedly difficult to differentiate between perception and cognition. However, traditionally, perception has been defined as the cognition of form, and we are therefore concerned with assessing the visual discrimination and recognition skills that are ordinarily basic to later competency in reading. The *Look-alikes* instrument samples the child's ability to match to a standard. Both open and closed figures are appropriate at the preschool level, and it is important that the child perceive a unit or form as separate from its background and discriminate among similar units and forms even under simple transformations. For example, in matching to a standard, the preschool child should be able to discriminate among such numerals as 6, 9, and 8 using 6 as the stimulus and among such lower case letters as b, n, and h using h as the stimulus. In *Look-alikes*, the child's ability to match series or groups of forms, objects, letters, and numerals is also assessed.

Equally important is the assessment of the child's discrimination and recognition skill in response to appropriate verbal labels. For example, in *Finding Letters and Numbers*, the child is required to recognize the verbal label *g* and to discriminate among the capital letters C, Q, and G

or, in response to the verbal label *a*, to discriminate among the lower case letters d, a, and o. For numerals, the child must respond to the verbal label "three" and discriminate among the numerals 5, 3, and 8 or, in response to the label "twelve," to discriminate among the numerals 12, 10, and 21. Knowledge of letter and number names and the ability to discriminate among similar open or closed and curved and/or straight forms generally precedes competencies developed through formal education.

The assessment of both receptive and productive skills and abilities is necessary for understanding, diagnosing, and prescribing for the child's educational needs. The production of open and closed forms can be discerned from the time of the child's first scribblings, but the preschool child should be able to reproduce or copy from a visually presented form in a controlled manner. In *Copy What You See*, this perceptual-motor coordination is assessed through the child's ability to reproduce such capital and lower case letters as X, P, f, and B and such numerals as 2, 7, and 5.

The Need for Information

Perhaps of greater importance than the need for information about specific understandings and skills for a given child is the need to provide teachers with information about age- and stage-appropriate levels and typical developmental progression from one level to the next. There is also an urgent need to provide information about parallel development across various cognitive areas. This will be possible through the array of measures being described in this symposium. The theoretical framework for each measure will be provided in its manual, and the various developmental sequences will be discussed. Summaries of group performance can be made available for system or statewide assessment. To help the teacher interpret the results of the child's performance across the areas sampled, individual profiles and sentence descriptions of performance will be provided.

Reference

Piaget, J., & Inhelder, B. *The psychology of the child.* New York: Basic Books, 1969.

PROBLEM SOLVING AND DIVERGENT PRODUCTION

Ruth B. Ekstrom
Educational Testing Service

Teaching children how to go about solving problems is considered by many to be the fundamental mission of education. It is argued that since it is impossible to predict what knowledge a child will need in this rapidly changing world, we should teach him how to solve problems, an education that will prepare him for a variety of situations and for subject matter yet to be discovered.

Closely related to problem-solving skills are those of divergent production, or creativity. The child who is limited in the realm of divergent production may be unable to solve a problem, not because she lacks the necessary reasoning skills but because she is unable to generate enough different hypotheses or possible alternatives.

For these reasons and because cognitive functioning and divergent production skills are often mentioned among the goals of preschool and kindergarten programs, we decided that it was important to include measures of both in the CIRCUS battery.

The Problem-solving Measure

Think It Through, the CIRCUS problem-solving measure, is designed to assess five essential abilities: 1) the ability to detect the problem, 2) the ability to define the problem, 3) the ability to use order and sequence in problem solving, 4) the ability to evaluate possible solutions, and 5) the ability to use classification skills in problem solving.

Detecting the Problem

We decided that a group of tasks asking the child to select incongruities would be the best technique for measuring the ability to be aware of a problem and to define it. These tasks involve the perception of missing parts (such as a table leg or the hands of a clock), physical impossibilities (such as water going uphill when poured), size incongruities (such as a door being too small in scale when compared with the rest of a house), and inappropriate usage (such as using a hairbrush as a toothbrush).

Because these problems always present two examples of similar "real world" possible situations along with the third incongruous example, it

is sometimes possible for the child to solve the problems in this section by a visual analysis of the three pictures and a comparison of their similarities and differences.

Defining the Problem

During the development of this test, we decided that measuring a child's ability to define a problem could best be accomplished by means of a relational-implicational reasoning type of task that would first require the child to develop the concept of a class from an array of objects and then ask him to select an object that does *not* belong to that class. Rational-implicational reasoning is one of three kinds of reasoning which nursery school and kindergarten objectives frequently mention.

The importance of concept formation in problem solving has been well documented. Maier (1936) showed nearly 40 years ago that preschool children can discover a principle and apply it to a new situation. It is important to note that recent research (Stevenson *et al.*, 1968) has shown that children's problem solving includes a variety of intellectual and motivational factors and is not simply the application of a single, general learning factor. Studies of learning among elementary school children (such as Duncanson, 1966) have also supported the existence of concept formation as a separate cognitive factor.

Concept formation is, of course, dependent upon previous experience: Solution of problems in this section is dependent on the facts and knowledge the child has available to him. For example, the child must be aware of the difference between animate and inanimate objects to reach the conclusion that a fire hydrant does not belong with a group that includes a horse, a dog, and an elephant.

Although the ultimate objective of these problems is to have the child select a single object that does not belong with the others, it is also important to stress, through these tasks, that objects can be looked at in many different ways and that flexibility and open-mindedness are essential to the concept-formation phase. The child who limits his attention to one or two characteristics of the object may be so restrictive in his analysis of the problem that he is unable to reach a solution.

Awareness of Order and Sequence

This section was included in *Think It Through* to emphasize that problems are involved not only with how large or small the object is, its

shape and function, but also with when the object is used and with the events which must precede and/or follow that use.

Tasks asking young children to remember sequences have appea ed in a variety of tests for this age group. However, this section of the CIRCUS problem-solving test is different because it is not primarily dependent upon short-term memory for sequence or order. Unlike the bead-stringing or block-tapping tests that appear in other test instruments, the *Think It Through* sequence items are primarily concerned with real-world events, such as drinking a bottle of pop, building a house, or going down a slide. The child who has observed or taken part in such activities can, of course, solve them by resorting to memory (as in the case of the incongruities items discussed earlier), but even without such knowledge he can reach the correct solution through logical analysis.

It is important to point out here that there are common elements among items across sections that are designed to alert the teacher to basic misunderstandings the child may have. For example, the child who is unaware of the effects of gravity may have difficulty in different test sections with items about pouring water, balancing on a seesaw, and going down a slide.

Sequence and order items are an elementary form of the more formal "if–then" thinking called postulation. Requiring an analytic-deductive approach, they are part of systematic reasoning, which is one of the three types of goals frequently mentioned in nursery school and kindergarten objectives.

In the fourth section of this test, we have built upon the idea that there are a variety of ways of looking at almost any problem and have developed items that represent another essential of problem solving–the ability to evaluate several possible solutions and to select from among them the one that is best.

Evaluating Possible Solutions

The ability to evaluate the degree of appropriateness of various responses (or solutions) and to think critically about the implications of each is related to systematic reasoning.

The problems in this portion of *Think It Through* are based on real-life situations. Each item consists of a stimulus picture accompanied by a short story that develops the problem and by three pictured solutions. While all of the solutions will solve the problem, one is clearly superior to the others. One item, for example, asks the child

to choose whether a lollypop, a bag of popcorn, or a ball of gum would be the easiest to share. Other items require the child to select the best way to reach a cookie jar on a high shelf, the best way to put up a poster, the best replacement for a broken shoelace, the most efficient way to give water to Ellie the elephant, and the safest way to retrieve a ball from the lion's cage.

The final section of *Think It Through* involves both one- and two-way classifications. (Classification is the third of the three types of reasoning frequently mentioned in nursery school and kindergarten objectives.) The child is required to recognize superordinate-subordinate and class membership relations based on common properties.

The simple classification problems deal with attributes such as shape, size, color, and function. These problems are similar to the concept-formation aspects of the problem-definition section. The two-way classifications pair these so that the child must consider both shape and color, size and shape, shape and function, and so on, simultaneously. Because of their earlier experiences in this test with similar concepts in less complex situations, even children who have had no previous experience with formal classification problems grasp the task readily.

The Divergent Production Measure

Because so many tests have a single, correct answer, we feel it crucial to include in CIRCUS at least one measure that will point out to teacher and pupil alike the importance of being able to produce a variety of *different* responses.

The concept of divergent thinking has been described by Guilford (1959) as thinking "in different directions, sometimes searching, sometimes seeking variety." According to Guilford, "the unique feature of divergent production is that a variety of responses is produced. The product is not completely determined by the given information." The problem-solving test includes tasks that Guilford would call "convergent production, cognition, and evaluation." The *Making Trees* test requires outcomes that are very different and might be associated with creativity.

In developing a divergent production measure, we wanted to stress ideational fluency and flexibility, both popular objectives in programs of early childhood education. Several researchers have explored verbal divergent production in preschool and kindergarten children with a uses-type task (Ward, 1968; Iscoe and Pierce-Jones, 1964; Biller *et al.*, 1969). However, we felt that the reluctance to verbalize, which is a

frequent problem in testing young children, and the vocabulary limitations, which are often found in children from culturally deprived or different backgrounds, made it desirable to attempt to develop a nonverbal divergent production task.

In *Making Trees*, the child is first presented with gummed labels of various geometric shapes and colors and is asked to make a tree. Later, she is shown her first product, is told that it was a "good" tree, and then is asked to make another as different as she can from her first.

Because this divergent production task was more experimental in nature than most of the other CIRCUS instruments, we have been conducting more extensive research on it. While we do not yet have the results from the norming administration of the entire battery, we do have considerable other information about it, which has been collected under the supervision and direction of Dr. William Ward.

Among the rather complex scores obtained on this task are the number of stickers used, the amount of elaboration (including objects not necessarily associated with a tree), the inclusion of extra trees, a rating of the aesthetic appeal of the construction, a rating of the appropriateness or "treeness" of the construction, a score for "minimal", or two-sticker, trees, a rating of the degree of unusualness of a child's tree as compared with those of his classmates, and a rating of the extent of difference between the two trees produced by the child.

An earlier attempt to include an originality score had to be dropped because originality and appropriateness ratings correlated about .9 in one small study. This was primarily because a number of children did a barely competent job in constructing trees and, as Dr. Ward has pointed out, "it is necessary to be competent before you can start to show originality."

Inter-judge reliability ratings (using coefficient alpha) for a single set of trees range from .72 for the aesthetic appeal rating to .90 for the appropriateness rating. The correlation of the ratings for two sets of trees from the same children range from .30 for the aesthetic appeal rating to .48 for appropriateness, suggesting that the children, indeed, made relatively different trees in the two administrations.

This promises to be a most exciting new measure, and we are looking forward to the availability of more information about it and the other measures in the CIRCUS battery when our normative data study becomes available.

References

Biller, H. B., Singer, D. L., & Fullerton, M. Sex-role development and creative potential. *Developmental Psychology*, 1969, 1, 291-296.

Duncanson, J. P. Learning and measured abilities. *Journal of Educational Psychology*, 1966, 57 (4), 220-229.

Iscoe, I. & Pierce-Jones, J. Divergent-thinking, age, and intelligence in white and Negro children. *Child Development*, 1964, 35, 785-798.

Guilford, J. P. Three faces of intellect. *American Psychologists*, 1959, 14, 469-479.

Maier, N. R. F. Reasoning in children. *Journal of Comparative Psychology*, 1936, 21, 357-366.

Stevenson, H. W., Hale, G. A., Klein, R. E., & Miller, L. K. Interrelations and correlates in children's learning and problem solving. *Monographs of the Society for Research in Child Development*, 1968, 33, No. 7.

Ward, W. C. Creativity in young children. *Child Development*, 1968, 39, 737-754.

THE CONTEXT OF ASSESSMENT AND THE ASSESSMENT OF CONTEXT

Samuel Messick
Educational Testing Service

Early childhood education is an extremely complicated *system*. It involves, at the very least, a set of complex, multifaceted organisms changing over time in interaction with diverse environmental influences. Furthermore, this system is composed of differentiated but overlapping subsystems that embrace the child, family, community, and various peer groups as well as the school, teachers, and programs. Since the concept of *system* implies a functioning whole whose various elements and subsystems are interdependent, it follows that the operation of one part of the system may interact with and produce unanticipated consequences in other parts of the system.

In attempting to measure any element or characteristic of such a system, we must assess the general context of interdependencies in order to take into account possible interactions of the characteristics measured with other aspects of the system—especially interactions among student, teacher, situation, and background characteristics. Otherwise we are at a loss to know how to generalize the measure and its meaning (or to limit its generalization) across student groups and across situations.

This relativity of inferences about measured characteristics to context has three major aspects:

1. Inferences about personal characteristics, particularly about competencies, should be relative to the context of environment, educational experiences, and programs to which the child has been exposed. When inferences about competency are drawn from test performance, it should make a difference whether or not the child has had an opportunity to learn the skills required by the task or whether the child (or his teachers or parents or peers) thought those skills were important or relevant.

2. Inferences about a particular characteristic or competency of a child should be relative to the context of his general personality and intellectual makeup, or at least to the salient features of that makeup. The child himself is a very complicated system of interdependencies, and one must anticipate that certain of his traits

and characteristics will influence or interfere with the assessment of other traits and characteristics.

3. Inferences about measured characteristics should be relative to the context of the measurement process *per se*, not just by taking into account critical objective features, such as whether the task was timed or untimed, but by tempering interpretations of test responses in light of the child's general style of reaction to the task, the tester, and the testing situation.

A comprehensive program of individual assessment should include provision for gauging these three major aspects of context, for if we are sensitive to the issues, even relatively primitive indicators of contextual interactions can have a profound influence on our interpretations. They can provide warning signals, for example, that certain generalizations may be unwarranted, that alternative hypotheses should be seriously entertained, or that additional measurement should be undertaken to clarify ambiguities.

Let us consider some strategies for the assessment of these three major aspects of context, as exemplified in the ETS CIRCUS approach to comprehensive assessment.

I. The Context of Environment, Programs, and Educational Experience

Environmental and program context is perhaps ideally assessed through direct observation using mutiple independent observers. It may also be conveniently and much less expensively assayed using indigenous, though biased, observers by means of a teacher questionnaire. Since teachers are prime agents in the educational context afforded the child, their biases are important to document in their own right, and a teacher questionnaire offers a ready means not only for eliciting teachers' descriptions of class and program characteristics, but also for appraising attitudes and viewpoints that might influence both their judgment and their teaching behavior.

Through this questionnaire mode, then, teachers are asked to describe the background of each child in their class in terms of age, sex, ethnic group membership, family occupational status, and previous educational experience; to describe the structure and setting of the classroom, the materials and facilities available along with the extent of their use, and the relative amounts of a variety of classroom activities;

and to characterize briefly the school or center of which the class is a part. In addition, the teachers are asked several questions about previous experience and education, job attitudes and preferences, educational viewpoints, and predilections for various educational techniques and objectives.

This direct questioning of teachers about their programs and preferences may draw their attention to gaps in desirable facilities and activities or to an underemphasis upon valuable techniques and objectives, which they may subsequently correct. This may be all to the good educationally, but we should be sensitive to the possibility that such a reactive approach to the assessment of context may be obtrusive and hence may change or distort the very context it is meant to assess. From a research standpoint, this is an interesting but possibly minor caveat. It points to one of many possible sources of reliable change in context and, given the general intractability of teacher behavior, not a very likely one at that. The more basic lesson it underscores should by now be a measurement commonplace—that the stability of any context, just like the reliability of its assessment, is an open empirical question and that the generalizability of a measure from one point in time to another requires recurrent response consistencies.

II. The Context of General Personality and Intellectual Makeup

The context of salient traits and characteristics comprising the child's effective personality and intellectual makeup is most directly assayed through a strategy of multivariate measurement and analysis. Rather than measuring a single characteristic in isolation, or even a collection of separate characteristics, one should assess and interpret multiple characteristics in relation to each other, using score or factor profiles or other forms of comparative and moderator analysis. Score interpretations should take into account evidence of interactive or moderator effects: A high score for a particular characteristic may have a different meaning or different implications for individuals scoring high as opposed to low on a second characteristic or for individuals displaying a particular pattern of scores over a set of characteristics. Thus, the educational implications of a low score on a general information test may be quite different for a child who achieved moderately well on a variety of measures of problem solving and cognitive functioning as opposed to a child who performed poorly on those tasks. Or a consistent pattern of moderate-to-low performances on cognitive tasks

might be interpreted somewhat differently if accompanied by an extremely low score for memory or recall as opposed to a moderate or average score.

In the construction of comprehensive assessment batteries for children, emphasis is understandably given to dimensions of intellectual attainment, cognitive functioning, and sometimes even creative process, for these are closely attuned to major educational and social objectives. Less time is typically allotted to the assessment of affective dimensions, not because they lack educational or social relevance, but primarily because of difficulty in developing valid and efficient measures in the affective domain. Yet it is just such affective variables as motivation and interest and coping that provide the critical personal context necessary for drawing valid inferences about process or competency from cognitive test performance.

Given the interpretative importance of these affective variables, an attempt has been made to assess them in the CIRCUS battery by turning once again to teachers' judgments. However, rather than asking teachers to make the kind of high-level inferences that are necessary to rate such characteristics as aggressiveness or achievement motivation, with all the inherent biases entailed by such value-laden content, the *Activities Inventory* asks them instead to rate each child in connection with a variety of activities. These activities, which include physical, motor, academic, language, role playing, fantasy, and artistic behaviors, are rated with respect to frequency of occurrence, degree of complexity, the creativity and imagination displayed, the amount of help or direction typically sought from adults, and the degree to which the child usually engages in the activity by himself. If these ratings are sufficiently discriminating across children and display individual variability across activities, then this *Activities Inventory* approach may provide serviceable measures of interests and of preferred or habitual coping styles in young children.

III. The Context of the Measurement Process

The context of the measurement process itself is most usefully assessed not so much by documenting objective characteristics of the tasks, the tester, and the situation as by recording the child's stylistic reactions to them. This is usually accomplished, following the lead of Hertzig *et al.* (1968), by means of direct tester or teacher observations of the child's stylistic responses to the cognitive demands or adaptive requirements of the measurement tasks. These ratings may be made separately for each

task, for a representative selection of tasks, or globally for the battery as a whole. The CIRCUS *Inventory of Test-Taking Behaviors* includes judgments of such aspects of the child's responsiveness as:

- the degree to which he asked for help
- refused or indicated reluctance to work on tasks
- expressed enjoyment or amusement over particular content
- indicated that he didn't know certain answers
- indicated a desire to stop
- appeared to respond at random
- appeared to weigh alternatives carefully
- spoke about or attended to unrelated objects or events

By relating stylistic consistencies in test responsiveness to patterns of test performance, the validity of test interpretation is likely to be improved, regardless of whether these response styles are transient and specific to particular tasks or situations or are more generally characteristic of the test-taking behavior of the subject.

From this discussion it would appear that the major approach to the assessment of context is observational, that it is difficult to avoid the intrusion of human judgment in the measurement process. Although at this stage of the art this may be true, it is not a critical issue to be emphasized here. The important point is not that the assessment of context is inherently observational, but that it is inherently analytical. Dimensions of context are important because their potential interactive and moderator effects may differentially influence individual behavior. Hence, the descriptive measurement of a variety of dimensions, however salient or pervasive, is not enough for a true assessment of context. Such descriptive measurement provides the setting and background, but *context* in education and psychology, as in discourse, is that part of the surround that illuminates or adds to the meaning of the subject. For that, the interactions and moderated relationships must be assessed or revealed analytically. For a comprehensive program of measurement to deal meaningfully with the assessment of context, then, it must include

provision for multivariate analysis and for the display, reporting, and interpretation of interactive and moderated relationships.

Reference

Hertzig, M. E., Birch, H. G., Thomas, A., & Mendeg, O. A. Class and ethnic differences in the responsiveness of preschool children to cognitive demands. *Monographs of the Society for Research in Child Development*, 1968, 33, Serial No. 117.

DISCUSSION

Boyd R. McCandless
Emory University

The first point that seems worth making is that, although many thoughtful Americans these days are not very happy with the role of private enterprise in the United States (nor altogether unhappy either), the present symposium appears to me to be a good illustration of a private organization's constructively leading the way in innovative test development with its staff members working as a team.

The overriding thrust of the excellent papers in this symposium seems to be a healthy going back to behaviorism. CIRCUS is a test of behavior—how children attack problems—rather than one more extension of testing experts into trait theory. The CIRCUS team members are not looking for an overriding single predictive score, such as the IQ, but rather are sampling behavior in a number of ways so as to guide teachers into diagnostic instruction. My own experience with inner-city teachers of poor black and white children has indicated that there are two major means by which psychologists can help teachers: through their knowledge of (1) principles of behavior management and (2) diagnostic teaching. CIRCUS seems to be a first-rate gambit for giving teachers guidelines for the latter.

CIRCUS is based on the difference, not deficit, hypothesis of children's development and learning. Children are *not* simply lower or higher than one another along a trait dimension of, for example, IQ. They are different. Some solve problems, talk, and think in different ways from others. The different ways are not necessarily better or worse than each other, although they may vary in efficiency. CIRCUS is designed to tap such differences, not to tell a teacher that one child is inferior to another. This is a valuable evaluation concept, of considerably more practical value than testing based on trait/deficit theory.

Implied throughout the development of the CIRCUS instruments is the advantage of a sequence-relevant rather than a normative or chronological age approach to the development of children. In the chronological age approach, we are told that "all six-year-olds are ready to read." We know that this is not true: Some children are ready to read at three or four years of age, others are not ready until much later. In the sequence-relevant approach, we look at where a child is—what skills he or she possesses—with the purpose of moving him on to the next higher stage of skill development and exercise. Such a point of view is a mix of

humanism (all people should be striving for the next higher level of integration and exercise of their abilities and potentials) and of cognitive developmental theory. The sequence-relevant approach is what Piaget calls "the American question." For example, can conservation of number be accelerated through teaching?

For skills that underlie learning, acceleration seems to this discussant to be desirable: If one knows how to conserve number or volume earlier, for example, he is equipped with learning strategies that open up new vistas that cannot but help him in school. While the family may be where the child "is formed," school is the arena in which he makes the first and perhaps the most lasting evaluations of his competence. CIRCUS seems to consist of a number of multifaceted instruments for assessing where a child is and, thus, is an interesting device for helping teachers move children along the sequence-relevant course toward more extensive and faster mastery of their learning environment.

Most of us agree that children who live in the United States culture today should be equipped to cope with it. Diagnostic techniques that help cultural mastery, then, must be good for children. CIRCUS, as it is being developed, seems to be a program that will foster competence among American school children. Thus, I am pleased to learn about it and support its rapid development and use.

DISCUSSION

Marshall S. Smith
National Institute of Education

The introduction of new measures to the early childhood field stimulates two reactions in me. First, I share with those who were not involved in the development the opportunity to applaud: Many of us are seeking the new test or set of tests which will make our research and evaluation tasks easier.

This enthusiasm is generally tempered, however, by the realization that most efforts to develop measures outside the conventional achievement areas have failed. As a consequence, my second reaction is to be skeptical about new instruments and to require considerable psychometric information about them before I make decisions about using or recommending them for use.

My task is, therefore, both easy and pleasant. I can applaud CIRCUS and make some general comment about its promise while at the same

time *suspending* complete judgment until data on the performance of these measures are available.

Leaving aside possible psychometric problems, let's look at the design and objectives of these instruments. By and large, I find myself in almost complete agreement with the approaches taken by the authors. Moreover, I feel that there is now a great need for a program like CIRCUS. Let me explain why, using two personal experiences.

A group of us recently completed a series of reports on Head Start Planned Variations, a large-scale field study which examined the effects on children of a number of different preschool curricula. One primary concern during the early planning of the study was to put together a battery of existing tests that would faithfully represent the variety of objectives suggested by different preschool curricula ranging from the Open Classroom type such as Bank Street to academically oriented curricula such as Englemen-Becker. Although we made great efforts to construct an appropriately comprehensive battery, we failed. Almost all of the chosen tests turned out to be close cousins of the standardized achievement test and many were extremely difficult to administer on a large-scale basis. Had the CIRCUS battery existed in a field-tested and reliable form, we might have been able to take two giant steps toward the solution of our problem:

1. First, it would have given us the means to begin to cope with the diversity of objectives represented by the preschool models. Having available a well-developed and at least partially comprehensive battery to choose tests from might have forced both the evaluators and the curriculum sponsors to estimate before the study began what the impact of the various curricula would be. Together we could have chosen the appropriate tests out of those available. This procedure might have given the curriculum sponsors a greater faith and investment in the evaluation. Since a number of the sponsors were naturally unhappy with the range of possible instruments and, therefore, with the choice of measures, they were also unhappy with the evaluation.

2. The use of a common format across the tests, the focus on ease of administration, and the emphasis on making the tests fun for children would have made the job of administering a battery of tests to 4,000 children faster, cheaper, and far less onerous. These are not trivial points—a single-battery administration in Head Start Planned Variations cost at least $150 and took roughly two hours.

> And those of you who have seen a child reduced to tears from the frustration and self-distaste induced by tests like the Stanford-Binet should appreciate any attempt to make assessment devices more pleasurable.

Not surprisingly, one of the main recommendations of our report suggested the construction of a battery with many of the same characteristics as CIRCUS.

A second recent experience of mine which reflects on the CIRCUS program involved a small group of people who were attempting to specify a set of objectives and a research agenda for the National Institute of Education in the area of linguistic communication. Our responsibilities included thinking about ways of making the teacher's job easier. Perhaps the most consistent recommendation we received—from researchers and teachers of all types—was to encourage the development of a theoretically coherent and easily administered set of instruments with which teachers could assess the competences—and thereby diagnose the problem areas—of individual children in their classrooms. From what I have heard today, CIRCUS may have taken us one-half the way to that goal. If the CIRCUS tests are as easily administered and as useful to teachers as the authors hope, we have a strong prototype for the reading competences battery. Then, of course, all we will need in order to suggest the appropriate competences to be tested is a valid and comprehensive theory of how children learn to read.

In summary, then, I applaud the authors of CIRCUS:

- first, for their attempt to produce a theoretically coherent and comprehensive set of assessment instruments for young children;
- second, for their focus on ease of administration and use by classroom teachers; and
- third, for their attempt to make the tests pleasant and relevant experiences for the children.

I await the results of the normative study: Good data suggesting that CIRCUS meets even minimal expectations will herald an important contribution to the field. Finally, I hope that research on CIRCUS will not cease with the normative study—in particular, I would like to see careful work done on the usefulness of the instruments for the classroom teacher.